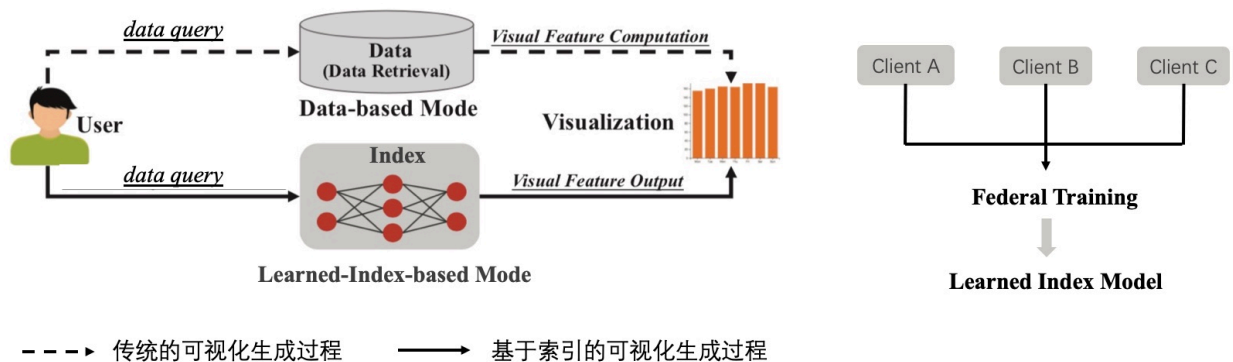


Weekly Report (2019.5.20-2019.5.26)

DONE

1. 联邦可视化

1. 参与林昊的 VIS for 联邦学习项目讨论，同时对联邦学习的一些细节有了更深一步的认识。中心 server 负责分发初始模型并且将 client 上交的参数进行平均并再次分发模型。client 负责训练模型。通过这种模式训练出更加完备的模型。
2. 联邦可视化想要从简单的统计做起，想到和之前做的 data cube 有些相同的地方，于是回顾了 nanocubes, hashedcubes, and gaussian cubes 中使用的方法，暂时没有想到如何去运用，先放弃这边的思路。
3. 看了李杰老师的投稿文章，学到了 learned index 这一个新概念，并阅读文章中提到的灵感来源 Jeffrey 的 The Case For Learned Index Structures。这篇文章的思想是，数据库的索引和机器学习里的预测模型其实有一些相似之处，比如 B 树是把 key 映射到一个有序数组中的某个位置，Hash 索引是把 key 映射到一个无序数组中的某个位置，bitmap 是把 key 映射成是一个布尔值（存在与否）。而这些事都可以拿预测模型来做，用模型去逼近原始数据，这样在查询的时候，就不需要原始数据，只要有这个模型就可以了。我觉得这个概念可以套用过来，框架如下图所示，在新的可视化框架下，数据分析师不需要接触到原始敏感数据既可以进行数据查询，并得到可视化图表。首先利用联邦学习训练出一个 learned index 模型，能够充分反映原始数据的数据分布特征。然后，用户在进行数据查询时，直接输入到模型中，模型给出一个结果，不需要遍历原始数据，有别于传统的可视化生成过程。通过这种方式可以打破数据孤岛，进行领域间的合作，尤其对于医疗、金融等领域。已经安排王智勇去实现了，数据使用的是rsatree中使用的 flight data。



2. Visevo 投稿项目

没有新的进展

3. 平安科技报告ppt

1. 在陈老师做的ppt框架下补充slides，包括阿拉丁系统的相关内容，联邦学习的简介，VIS for 联邦学习以及联邦可视化。

4. Chinavis审稿

完成Chinavis的审稿。

5. 准备深圳之行

与新加坡同学进行沟通，确定往返时间，并一起预定机票。

小结

工作日工作时长 $9+10+9+9+9$ ，周末 $12h$ ，总时长约 $58h$ 。

PLAN

短期计划（一周）

1. 按照现在的联邦可视化思路做出点东西

中期计划

1. 完成 Visevo 投稿项目
2. 调研联邦学习
3. 准备金融可视化的综述

长期计划

1. 学习更多机器学习、数据挖掘相关的算法。
2. 在项目中锻炼自己的思考能力与代码能力。